

MERCATUS CENTER

GEORGE MASON UNIVERSITY

Government Accountability Project
Working Paper

**An Analysis of the Office of Management and
Budget's Program Assessment Rating Tool
(PART) for Fiscal Year 2008**

EILEEN NORCROSS
JOSEPH ADAMSON

Introduction

With the release of the Bush Administration's proposed budget for FY 2008, the Office of Management and Budget (OMB) has completed assessing 973 programs representing 96 percent of the federal budget with the Program Assessment Rating Tool (PART). PART was initiated with the FY 2004 budget in an effort to encourage agencies to report on programs goals, and results, and develop performance measures in order to help inform funding decisions.

PART is one element of the Bush Administration's Budget and Performance Integration initiative to link performance information to budgeting decisions – known as performance budgeting. A performance budget is, “an integrated annual performance plan and annual budget that shows the relationship between funding levels and expected results. It indicates that a goal or set of goals should be achieved at a given level of spending.”¹ Some form of performance budgeting has been adopted by nearly all 50 states, as well as internationally.² Though the extent to which it is applied varies.

OMB presents PART as the Executive's attempt to motivate agencies to comply with the Government Performance and Results Act (GPRA), passed by Congress in 1993. GPRA requires agencies to articulate the results they seek to accomplish for citizens, state how they will measure those results, and report annually on the measures. Agencies are further required by GPRA to produce strategic plans, annual reports, and performance reports.

These efforts to produce results information in agencies are not new. Several presidential administrations have offered their own versions of budgetary reform: notably, the Johnson' Administration's Planning Programming Budgeting System and Zero-Based Budgeting in the Carter Administration.³ These reforms—all different in design—share on element: they all try to address problem inherent in public sector budgeting—that is, where prices and profits don't exist how do policymakers know how to allocate resources among competing activities?⁴

Because policymakers “lack a budgetary theory”⁵ resource allocation decisions are more often made incrementally⁶ rather than based on results information. PART is the latest attempt at resolving this information deficit by systematically collecting data on program performance.

This paper does not assess how well PART fills the informational void, or if it is a consistently applied, or reliable evaluation tool. Nor can we determine in this analysis if

¹ John Mercer, *Performance Based Budgeting for Federal Agencies*, AMS, Fairfax, 2002 p. 2.

² John B. Gilmour and David E. Lewis, “Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores”, *Public Administration Review*, (Sept/Oct 2006): 742.

³ Jonathan Breul, *Presidential Management Reform Initiatives* (in *Getting Results: A Guide for Federal Leaders and Managers* ed. Kathryn Newcomer and Barry White) (2005): 59.

⁴ Gilmour and Lewis, “Assessing Performance Budgeting at OMB” (2005): 169.

⁵ *Ibid.*, The authors note the “lack of a budgetary theory was first identified by V.O. Key, see “V.O. Key, The Lack of a Budgetary Theory. *American Political Science Review* (1949): 34, 1137-44.

⁶ Aaron Wildavsky, *The Politics of Budgetary Reform*, Little Brown and Company, 3rd ed. 1979.

PART is actually informing budgetary decisions. Instead, we analyze the results of PART to see how agencies have been rated over the five year period. We examine how much of agency budgets fall into different ratings categories, how ratings have changed for programs assessed multiple times, and how different types of programs fare in the PART ratings.

I. PART’s Methodology and Application

As of FY 2008, OMB has assessed 973 programs, representing 96 percent of federal spending. During the past five years, upon an agency’s request, 186 programs have been assessed more than once.

The PART tool is a series of 25 to 30 *Yes/No* questions submitted to program managers. The questionnaire includes four sections—each weighted differently—dealing with a different aspect of program performance: purpose and design (20 percent), strategic planning (10 percent), program management (20 percent) and results/accountability (50 percent). The individual assessments for each program are posted on OMB’s website: ExpectMore.gov.⁷

Table 1 displays the questions asked in a PART assessment.

Table 1: PART questions

Section 1: Program Purpose and Design	<i>Weight = 20 percent</i>
1.1	Is the program purpose clear?
1.2	Does the program address a specific and existing problem, interest, or need
1.3	Is the program designed so that it is not redundant or duplicative of any Federal, state, local or private effort?
1.4	Is the program free of any major design flaws that would limit the program’s effectiveness or efficiency?
1.5	Is the program effectively targeted, so program resources reach intended beneficiaries and/or otherwise address the program’s purpose directly?
Section 2: Strategic Planning	<i>Weight = 10 percent</i>
2.1	Does the program have a limited number of specific long-term performance measures that focus on outcomes and meaningfully reflect the purpose of the program?
2.2	Does the program have ambitious targets and timeframes for its long-term measures?
2.3	Does the program have a limited number of specific annual performance measures that demonstrate progress toward achieving the program’s long-term measures?
2.4	Does the program have baselines and ambitious targets and timeframes for its annual measures?

⁷ See, <http://www.expectmore.gov>

2.5	Do all partners (including grantees, sub-grantees, contractors, cost-sharing partners, etc.) commit to and work toward the annual and/or long-term goals of the program?
2.6	Are independent and quality evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?
2.7	Are budget requests explicitly tied to accomplishment of the annual and long-term performance goals, and are the resource needs presented in a complete and transparent manner in the program's budget?
2.8	Has the program taken meaningful steps to correct its strategic planning deficiencies?
Section 3: Program Management	<i>Weight = 20 percent</i>
3.1	Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?
3.2	Are Federal managers and program partners (grantees, subgrantees, contractors, cost-sharing partners etc.) held accountable for cost, schedule and performance results?
3.3	Are all funds (Federal and partners') obligated in a timely manner and spent for the intended purpose?
3.4	Does the program have procedures (e.g., competitive sourcing/cost comparisons, IT improvements, appropriate incentives) to measure and achieve efficiencies and cost effectiveness in program execution?
3.5	Does the program collaborate and coordinate effectively with related programs?
3.6	Does the program use strong financial management practices?
3.7	Has the program taken meaningful steps to address its management deficiencies?
Section 4: Program Results/Accountability	<i>Weight = 50 percent</i>
4.1	Has the program demonstrated adequate progress in achieving its long-term outcome performance goals?
4.2	Does the program (including program partners) achieve its annual performance goals?
4.3	Does the program demonstrate improved efficiencies or cost effectiveness in achieving program performance goals each year?
4.4	Does the performance of this program compare favorably to other programs, including government, private, etc., that have similar purpose and goals?
4.5	Do independent and quality evaluations of this program indicate that the program is effective and achieving results?

The results/accountability section (section four) of PART receives the greatest weight. This section is designed to determine if the program has met or achieved efficiencies in its long-term performance goals and how the program's performance compares with

similar programs. This section may include performance evidence from independent evaluations, in addition to performance measures. Though much emphasis is given to the results section in terms of the weight in grading, Gilmour and Lewis find evidence that the process-oriented sections may be more influential in determining the overall score for a program. The authors suggest this may be because since many programs lack results information, budget examiners are forced to use information from the other three sections as proxies for performance.⁸

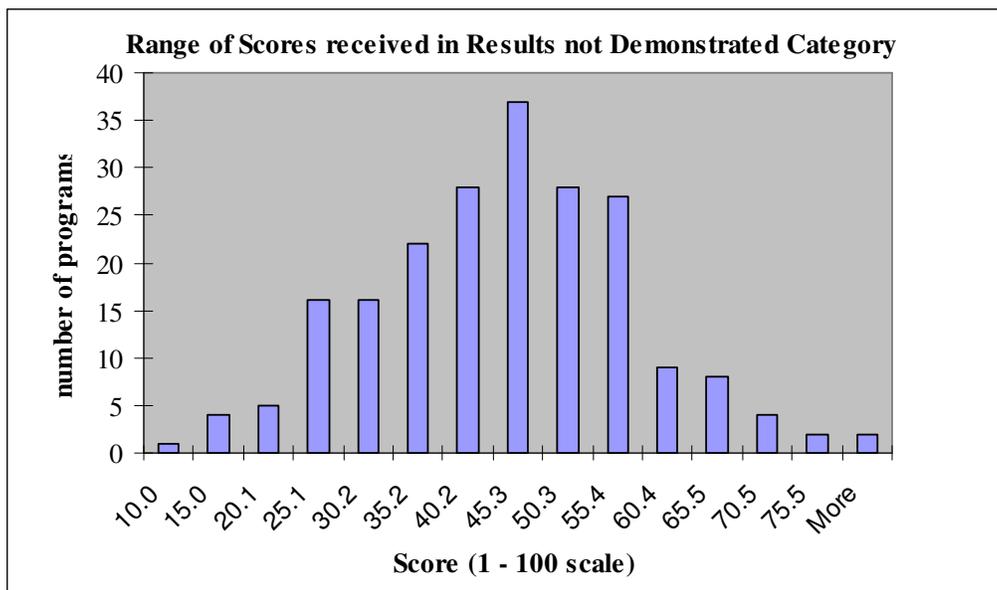
For each question the program receives a numerical score. The grades are summed for each section and multiplied by its respective weight. (purpose and design = 20 percent, strategic planning = 10 percent, management = 20 percent and results = 50 percent). The total scores for each of the four sections are summed to form the overall numerical grade.

A program may receive one of five qualitative ratings based on their percent grades: ineffective (0 to 49), adequate (50 to 69), moderately effective (70 to 84) and effective (85 to 100). A results not demonstrated rating may be awarded if the program lacks sufficient outcome measures, as indicated in questions 2.1, 2.2 and 2.3.

It is possible to receive a high numerical score in all other questions on PART and still get a results not demonstrated rating.

The distribution below shows the range of percent scores for programs receiving a results not demonstrated rating.

Chart 1: Distribution of percentage scores for results not demonstrated rating



⁸ Gilmour and Lewis, "Assessing Performance Budgeting at OMB" (2005): 185.

In the FY 2008 budget, two programs received a numerical score of 80.5 the National Credit Union Administration and Commodities Futures Trading Commission. This score would normally translate into a moderately effective rating.

The Commodities Futures Trading Commission received 100 percent in purpose and design, and program management, 71 percent in planning and 67 percent in results. However, it got a zero on question 2.1 and 2.2 which ask if the program has sufficient outcome measures—sufficient to warrant a results not demonstrated rating.

II. How PART has rated programs cumulatively

With the release of the FY 2008 budget, nearly all government programs have been evaluated at least once in PART. Over five years of PART, the number of programs with ineffective or results not demonstrated ratings has decreased, from five percent to three percent, and from 50 percent to 21 percent, respectively. Programs with effective ratings have increased dramatically from six percent to 17 percent. Similarly, those rated moderately effective have risen from 24 percent to 31 percent, and those with adequate ratings also increased from 15 percent to 28 percent.

This could be due to a few factors. a) Programs are getting better at “taking the test”, furnishing information to satisfy the examiners. b) OMB’s evaluations are getting more accurate or less accurate, that is, OMB is relaxing its criteria, or c) agencies are developing better performance measures and providing better information, that is, the quality of performance information has improved, or d) Agencies are improving in their actual performance because they are being measured – leading agencies to make qualitative changes in how their programs are designed and managed, resulting in improved outcomes.

The steady increase in scores shows that very few programs fall in their ratings. This gives credence to the theory that over time, agencies tend to do better on their assessments because they learn to answer the questions to the liking of the evaluators. It could also mean that programs evaluated for the first time in the most recent budget cycle (FY08) had several years to prepare for their evaluation and develop performance measures to satisfy PART’s requirements.

Chart 2: Cumulative program results by ratings category

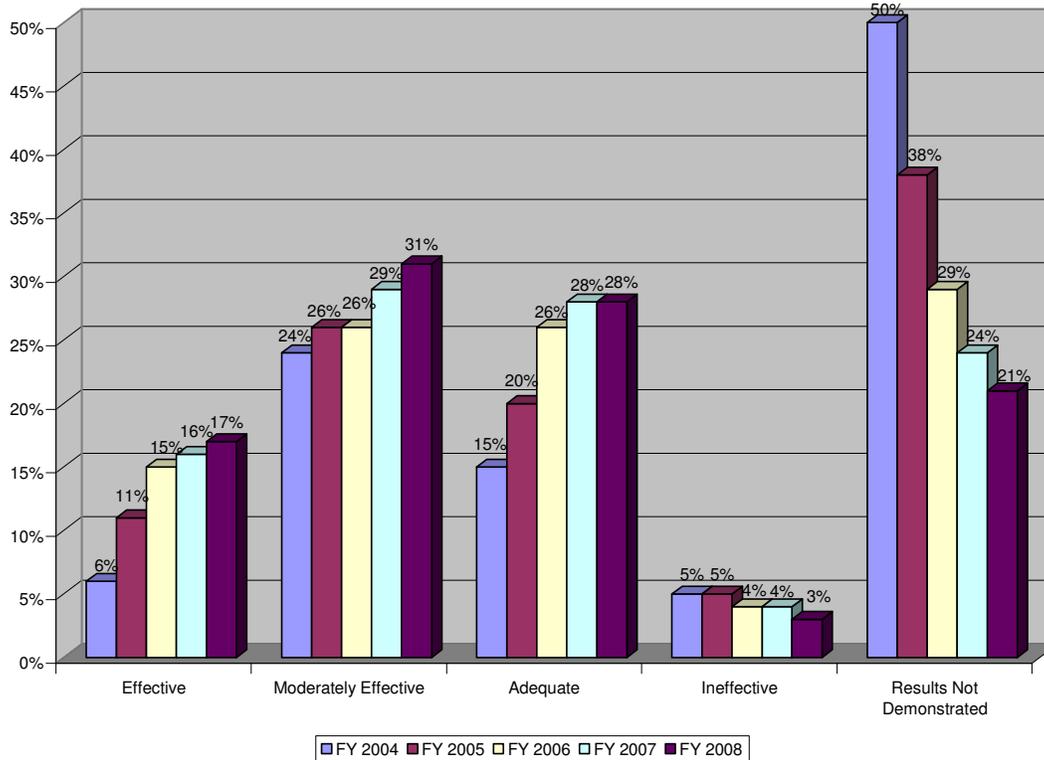


Table 2: How have reassessed programs performed?

Remained Results Not Demonstrated (RND)	12
RND to Ineffective	3
RND to Adequate	55
RND to Moderately Effective	37
RND to Effective	20
Ineffective to RND	1
Remained Ineffective	1
Ineffective to Adequate	3
Remained Adequate	10
Adequate to Moderately Effective	7
Adequate to Effective	2
Moderately Effective to Adequate	5
Remained Moderately Effective	15
Moderately Effective to Effective	11
Remained Effective	4

Every program evaluated in PART is supposed to be reassessed once every five years. An agency may request an earlier reassessment if they believe a program has made significant improvement. To date 186 programs have requested and received reassessment, of these, 166 have been rated twice. Nineteen have been rated three times. One program, Missile Defense, has been rated four times.

Some agencies aggressively seek reassessment. Agencies requesting assessments for ten or more programs include HHS (10), GSA (11), Education (15), EPA (17), State (18), Energy (19) and Agriculture (27).

State benefited the most, with fourteen of its seventeen reassessments increasing the programs' ratings to effective. One possible reason for State's exceptional performance is offered by Gilmour. He notes that State Department senior officials took PART very seriously from the start. Former Secretary Colin Powell "was deeply concerned with management of the department" and also appointed good managers as assistant secretaries.⁹ In addition, the agency's Bureau of Resource Management oversaw and organized the PART process, including 'after action' reviews of PART assessments to see what went right or wrong.¹⁰

Most of Agriculture's reassessments moved programs out of results not demonstrated to an adequate or moderately effective rating.

Overall, most programs tend to show improvements in a subsequent evaluation. Of 127 programs initially rated results not demonstrated 12 retained the rating, and three moved to ineffective (Community Service Employment for Older Americans, Youth Activities and Trade Adjustment Assistance, all housed in the Department of Labor)

The remaining results not demonstrated programs improved: 55 increased their score to adequate, 37 moved to moderately effective and 20 received an effective rating upon reevaluation.

About 88 percent of programs initially rated results not demonstrated subsequently improved their scores. These programs are 'self-selected' and focused on improving their PART evaluations. OMB indicates programs should be certain they have sufficient evidence to increase their ratings. That means either these programs are 'high achievers' taking the PART evaluation seriously – gathering the right evidence, developing good performance measures, and establishing good working relationships with their budget examiners. Or it could indicate that it is relatively easy to move from results not demonstrated by furnishing a 'bare minimum' of information.

The greatest number of programs jumping from results not demonstrated to effective are in the State Department. Other examples of programs moving from results not

⁹ John B. Gilmour, "Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrating Budget and Performance" IBM Center for the Business of Government: (2006): 15.

¹⁰ Ibid.,16.

demonstrated to effective upon reassessment include National Assessments, Adult Education State Grants, and the National Center for Educational Statistics all in the Department of Education, and three Energy Department programs: Nuclear Physics, Biological and Environmental Research, Basic Energy Sciences, Elimination of Plutonium-Grade Weapons.

Programs in other ratings categories do not show the same rate of improvement. Of the 19 programs initially rated adequate, more than half retained the adequate rating. Seven moved to moderately effective, and two more – both State Department programs - were rated effective upon reassessment (Refugees to Israel and Refugee Admissions to the U.S.).

Only six programs dropped in their ratings upon reassessment. The Department of Education's Safe and Drug Free Schools Grant moved from Ineffective to Results not Demonstrated. Five dropped from moderately effective to adequate: Energy's Nuclear Power 2010 program, the State Children's Health Insurance Program in HHS, Federal Pell grants in the Department of Education, Missile Defense in the Department of Defense, and NASA's Space Shuttle program.

Some programs show no improvement. Of the eight programs reevaluated in the Commerce Department, five retained their original scores.

Overall, fifteen programs remained moderately effective, while eleven others increased their rating to effective. Four programs remained effective upon reassessment: Energy Conservation in the Department of Defense, the Bureau of Economic Analysis in the Commerce Department, the State Department's Demining program, and NASA's Solar System Exploration program.

Ratings according to program category:

PART classifies programs according to seven categories, reflecting the funding type of the program.

- 1) Block/Formula Grants – Programs that provide funds to state, local, and tribal governments and other entities by formula block grant. Examples include the Community Development Block Grant, Medicaid, and Preventive Health and Health Services Block Grant.
- 2) Capital Acquisition – Programs that achieve their goals through development and acquisition of capital assets (such as land, structures, equipment, and intellectual property) or the purchase of services (such as maintenance, and information technology.) Examples: Navy Shipbuilding and General Service Administration – Real Property Leasing.
- 3) Competitive Grants – Programs that provide funds to state, local and tribal governments, organizations, individuals and other entities through a competitive process. Examples include Empowerment Zones and Safe Schools/Health Students.

- 4) Credit – Programs that provide support through loans, loan guarantees, and direct credit. Examples: SBA’s 7a loan program, Federal Housing Authority Multifamily Development.
- 5) Direct Federal – Programs where services are provided primarily by employees of the federal government. Examples: Federal Mint, Diplomatic and Consular Services, National Wildlife Refuge System.
- 6) Regulatory Based – Programs that accomplish their mission through rulemaking that implements, interprets, or prescribes law or policy, or describes procedure or practice requirements. Example: EPA’s Oil Spill Control and the Nuclear Regulatory Commission – Nuclear Reactors.
- 7) Research and Development – Programs that focus on knowledge creation or its application to the creation of systems, methods, materials, or technologies. Examples: Water Quality Research and Solar System Exploration
- 8) Mixed – Programs that combine elements from two or more categories (e.g. a research and development program that uses grants to fund research)

Examining PART ratings by type of program indicate that the funding mechanism matters and that there may be systematic reasons why certain types of programs do better than others. Block grants and competitive grants have the largest percentage—each with 32 percent—of programs rated results not demonstrated. This is a slight improvement from last year, when both categories had 36 percent of programs rated results not demonstrated. As with last year, the next lowest program types with a large proportion of results not demonstrated rating include Capital Assets (19 percent), Direct Federal (18 percent) and Credit programs (15 percent).

R&D programs perform the best in the FY2008 PART with 31 percent of all programs receiving an effective rating, followed by Regulatory programs with 23 percent. Direct Federal and Capital Assets perform relatively well with 20 percent and 19 percent of their programs rated effective, respectively.

These results indicate that a program’s funding mechanism may have implications for how the agency monitors and measures the program’s performance. For example, block grants are federal grants designed to be flexibly applied on the local level. Reporting requirements are reduced in order to minimize administrative burdens on grantees. This can result in less consistent data collection and monitoring, leaving little evidence of program performance for evaluators to use. Block grants fare worst in the results section of the PART, on average receiving a score of 36 percent (out of 100). The average on the results section for all programs is 49 percent. OMB suggests that block grants’ underperformance in the results category may be due to the fact that block grant funds are often awarded according to a formula, or based on population, not needs.

R&D programs, by contrast, tend to receive on average, much higher scores in the results section at 63 percent. Regulatory programs also receive higher marks on average, in results with a score of 51 percent in this section of the PART.

In the *Analytical Perspectives* section of the FY2008 budget, OMB states that measuring R&D programs for results presents inherent challenges since research can lead to unanticipated results. These programs are evaluated according to “R&D Investment Criteria” developed by the administration that takes these measurement constraints into account. Specifically designed questions for R&D are:

Table 3

Purpose and Design: R&D specific questions	
RD 1.	Does the program effectively articulate potential public benefits?
RD 2.	If an industry-related problem, can the program explain how the market fails to motivate private investment?
Strategic Planning: R&D specific questions	
RD 1.	Is evaluation of the program’s continuing relevance to mission, fields of science, and other “customer needs conducted on a regular basis?
RD 2.	Has the program identified clear priorities?
Program Management: R&D specific questions	
RD 1.	Does the program allocate funds through a competitive, merit-based process, or, if not, does it justify funding methods and document how it is maintained?
RD 2.	Does competition encourage the participation of new/first-time performers through a fair and open application process
RD 3.	Does the program adequately define appropriate termination points and other decision points?
RD 4.	If program includes technology development or construction or operation of a facility, does the program clearly define deliverables, capability/performance characteristics, and appropriate, credible cost and schedule goals?
Program Results: R&D specific questions	
RD 1.	If the program includes construction of a facility, were program goals achieved within budgeted costs and established schedules?

R&D programs are inherently difficult to measure their high performance in PART is worth considering. It may mean that these programs have been effective at measuring their performance, or that OMB grades R&D programs on a more lax basis because they are difficult to measure.

Regulatory programs also fare well in PART with high percentages of programs rated moderately effective (29 percent) and effective (23 percent). Regulatory programs are also subject to questions specific to their activities, shown below.

Table 4

Planning and Design: Regulatory specific questions	
Reg 1.	Are all regulations issued by the program/agency necessary to meet the stated goals of the program, and do all regulations clearly indicate how the rules contribute to achievement of the goals?
Management: Regulatory specific questions	
Reg 1	Did the program seek and take into account the views of affected parties including state, local and tribal governments and small businesses in drafting significant regulations?
Reg 2.	Did the program prepare, where appropriate, a Regulatory Impact Analysis (RIA) that comports with OMB’s economic analysis guidelines and have these RIA analyses and supporting science and economic data been subjected to external peer review, as appropriate, by qualified specialists?
Reg 3.	Does the program systematically review its current regulations to ensure consistency among all regulations in accomplishing program goals?
Reg 4.	In developing new regulations, are incremental societal costs and benefits compared?
Reg 5.	Did the regulatory changes to the program maximize net benefits?
Reg 6.	Does the program impose the least burden, to the extent practicable, on regulated entities, taking into account the costs of cumulative final regulations?
Results – Regulatory specific questions	
Reg 1.	Were programmatic goals (and benefits) achieved at the least incremental societal cost and did the program maximize net benefits?

It may be that regulatory programs have objectives that are easier to identify and measure: such as reductions in incidences of pollution, or enforcement activities. Since executive regulatory agencies (those with Chief Financial Officers) are required to perform cost-benefit analyses this information may help in completing the PART assessment. The extent to which regulatory agencies collect and use cost-benefit information varies.

The Federal Communications Commission is an independent regulatory agency. Five of the six programs assessed under PART received results not demonstrated scores. The individual assessments indicate that FCC is not measuring its activities or using performance data in a meaningful way. The Universal Service Fund programs receive very weak scores in management. They score very poorly in the planning section, and get zeros in results.

By contrast the Nuclear Regulatory Commission (NRC), an executive regulatory agency, is required to do cost-benefit analyses. NRC scores very well on PART. Four of its five programs are rated effective.

On questions designed specifically for R&D, the Nuclear Regulatory Commission programs do very well. In particular, NRC programs perform very highly on regulatory-specific questions. The individual assessments state that the agency conducts regulatory impact analyses on all its regulatory programs and that these analyses show a net benefit. It may be that the executive regulatory agencies receive greater scrutiny of their activities through the cost-benefit and other reporting requirements and that this results in higher PART scores.

Direct Federal programs also have high percentages of moderately effective (31 percent) and effective (20 percent) ratings.

Table 5: FY08 PART ratings according to program category

	RND	Ineffective	Adequate	Moderately Effective	Effective
Block Grant (155)	32%	5%	31%	24%	8%
Capital Assets (86)	19%	2%	27%	34%	19%
Competitive Grant (172)	32%	5%	30%	24%	9%
Credit Program (34)	15%	3%	38%	35%	9%
Direct Federal (342)	18%	1%	31%	31%	20%
Mixed (4)	25%	0%	0%	50%	25%
Regulatory (69)	22%	0%	26%	29%	23%
R&D (115)	6%	3%	15%	45%	31%

PART Ratings by program topic

For the past two years, OMB has assigned a topic to each PARTed program, based on sub-categories of the federal budget codes.

Table 6: Programs rated by topic

Topic	RND	Ineffective	Adequate	Moderately Effective	Effective
Agriculture (88)	15%	1%	39%	38%	8%
Business and Commerce (97)	21%	3%	30%	30%	16%
Community					

and Regional Development (56)	25%	7%	39%	20%	9%
Disaster Relief (25)	12%	4%	28%	28%	28%
Education (130)	42%	4%	25%	12%	16%
Energy (79)	9%	3%	13%	44%	44%
Foreign Affairs (95)	9%	0	25%	29%	36%
Government Administration (81)	17%	1%	33%	25%	23%
Health and Well-Being (180)	22%	4%	32%	31%	11%
Housing (43)	28%	9%	23%	30%	9%
Law Enforcement (82)	22%	1%	30%	30%	16%
National Security (134)	14%	0	17%	34%	35%
Natural Resources and the Environment (174)	16%	2%	41%	31%	31%
Science and Space (55)	16%	0	15%	29%	40%
Training and Employment (50)	24%	8%	38%	26%	4%
Transportation (61)	28%	2%	10%	48%	13%
Veterans Benefits (12)	25%	0	25%	50%	0

Certain areas of government activity, across agencies, tend to receive better ratings than others. Education programs continue to contain the highest percentage of results not demonstrated programs at 42 percent of all assessed. More than one quarter of housing, veterans' benefits, community development and transportation programs are results not demonstrated. High performing areas of the budget include science and space programs (40 percent effective), energy (44 percent effective) and foreign affairs (36 percent effective). Some topic areas match very closely to individual agencies while other topic areas are more dispersed across several agencies.

Programs rated by agency

Some agencies do better in PART than others. This is likely due to a variety of factors. Gilmour notes there may be systematic differences that account for the variation.¹¹ State, Treasury, Transportation, GSA, Energy, and Defense have done very well on PART. As noted, State took the assessment process very seriously from the start and aggressively pursued reassessment when ratings of particular programs fell short. Energy and GSA also pressed for re-evaluation of a large number of programs, indicating commitment to improving their scores.

GPRA requirements may have some effect on how agencies perform on PART. Executive level agencies are required to report on their performance, planning and program purpose and design. This information may help some agencies more effectively answer questions on their PART assessments.

Another reason for some agencies exceptional performance in PART is that some have high concentrations of a particular type of program. R&D programs and regulatory programs tend to be rated more highly, on average, than other kinds of programs, such as block grants. An agency with a high concentration of R&D programs is likely to fare better than one that administers a large number of block or competitive grants.

All thirteen of the National Science Foundation's programs are rated effective, representing \$6.5 billion. These are all research and development programs, which are evaluated slightly differently by OMB. Extra questions focus on whether the program prioritizes funding decisions, and if it assesses the potential benefits of its projects. OMB acknowledges that R&D programs present their own evaluations challenges – research often leads to unpredictable outcomes. NASA is another R&D-based agency with 100 percent of its programs rated effective.

The Nuclear Regulatory Commission is a regulatory-based agency. Four of its five programs also received effective ratings. One was rated moderately effective. On average, regulatory programs fare well in PART. This might be because regulatory programs are charged with relatively straightforward tasks that can be easily measured – issuing licenses, and enforcing codes.

Education programs have consistently underperformed in PART, with more than half receiving results not demonstrated ratings. Gilmour notes that this may be a part of a motivational strategy inside the agency. Education recognizes that it has many programs that are fundamentally flawed, 'burdened with ill-conceived, poorly designed programs.'¹² PART helps the agency highlight these deficiencies.

¹¹ Gilmour (2006): 15.

¹² John B. Gilmour "Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrated Budget and Performance, IBM Center for the Business of Government (2006): 16.

Other low-performers with high percentages of results not demonstrated programs include DHS (42 percent), Army Corps of Engineers (37 percent), Interior (30 percent), Veterans Affairs (30 percent), HUD (27 percent) and HHS (24 percent).

Agency performance in PART can be seen in a few ways. We can look at the percent of the agency’s programs according to the ratings to get a sense of overall performance in PART. As Chart 3 and Table 7 show, some agencies do quite well. The National Science Foundation, Social Security Administration, the Nuclear Regulatory Commission and Department of Transportation, Energy, GSA and State, all receive high ratings with a large majority of their programs rated either effective or moderately effective. By contrast the Department of Education, Department of Homeland Security, Veterans Affairs and Interior do poorly in PART with a large percentage of their programs rated results not demonstrated. HUD has a high percentage of programs rated ineffective (13.3 percent)

Chart 3: Agency performance in PART

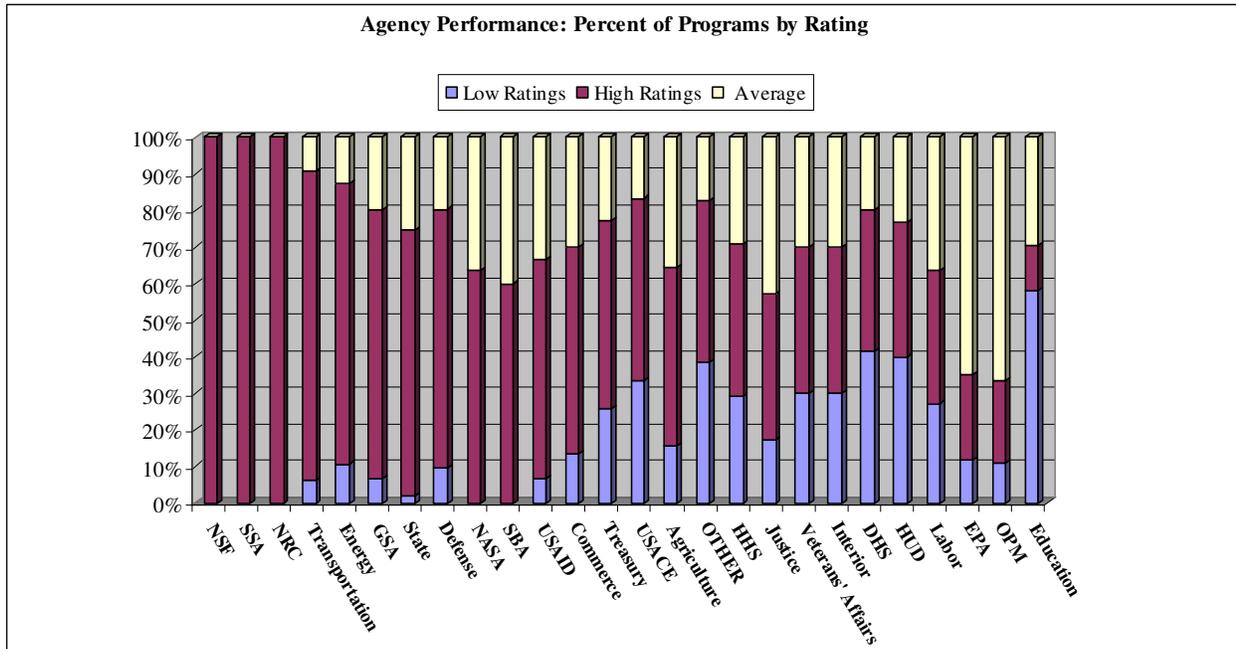


Table 7: Percent of programs rated in each agency according to PART scores

Percent of programs						
AGENCY	Results not demonstrated	Ineffective	Adequate	Moderately Effective	Effective	Total
Agriculture	15.5%	0.0%	35.7%	39.3%	9.5%	100.0%
Commerce	13.3%	0.0%	30.0%	40.0%	16.7%	100.0%
Defense	9.8%	0.0%	19.6%	39.2%	31.4%	100.0%
Education	53.4%	4.5%	29.5%	8.0%	4.5%	100.0%
Energy	7.1%	3.6%	12.5%	53.6%	23.2%	100.0%
HHS	23.9%	5.3%	29.2%	31.0%	10.6%	100.0%
DHS	41.5%	0.0%	20.0%	24.6%	13.8%	100.0%
HUD	26.7%	13.3%	23.3%	26.7%	10.0%	100.0%
Justice	14.3%	2.9%	42.9%	28.6%	11.4%	100.0%
Labor	15.2%	12.1%	36.4%	33.3%	3.0%	100.0%
State	2.1%	0.0%	25.5%	17.0%	55.3%	100.0%
Interior	30.1%	0.0%	30.1%	28.8%	11.0%	100.0%
Treasury	22.9%	2.9%	22.9%	17.1%	34.3%	100.0%
Transportation	3.1%	3.1%	9.4%	62.5%	21.9%	100.0%
Veterans' Affairs	30.0%	0.0%	30.0%	40.0%	0.0%	100.0%
EPA	5.9%	5.9%	64.7%	23.5%	0.0%	100.0%
NASA	0.0%	0.0%	36.4%	45.5%	18.2%	100.0%
NSF	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
SBA	0.0%	0.0%	40.0%	50.0%	10.0%	100.0%
SSA	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
GSA	6.7%	0.0%	20.0%	46.7%	26.7%	100.0%
NRC	0.0%	0.0%	0.0%	20.0%	80.0%	100.0%
USAID	6.7%	0.0%	33.3%	60.0%	0.0%	100.0%
OPM	11.1%	0.0%	66.7%	11.1%	11.1%	100.0%
USACE	33.3%	0.0%	16.7%	50.0%	0.0%	100.0%
OTHER	36.5%	1.9%	17.3%	19.2%	25.0%	100.0%

This is only one way of looking at agency performance. Another way to consider agency scores is to consider what these ratings represent in terms of their percentage of the agency's budget.

Agency program ratings as a percent of FY 2006 appropriations

If an agency's budget represents an investment in particular policy outcomes and PART scores measure the effectiveness of activities designed to meet those policy objectives, then it is meaningful to ask: how do PART scores match up with agency budgets. That is, what percentage of any given agency's budget is effective or ineffective?¹³

The U.S. Army Corp of Engineers and Veterans Affairs have the largest portion of their budgets rated results not demonstrated at 58 percent and 51 percent respectively. While both agencies have 30 percent of their total programs rated results not demonstrated this represents a much larger overall portion of their budgets. Interior also has a large portion of its budget rated results not demonstrated at 42.9 percent.

Veterans Affairs has an annual budget of \$71 billion. One of its largest programs, Disability Compensation, with budgetary authority of \$30.9 billion in FY 2007 is rated results not demonstrated. Veterans Home Loans (\$6.5 billion) and the Montgomery GI Bill – Education Benefits (\$2.8 billion) are also results not demonstrated. Taken together this accounts for 57 percent of VA's funding in FY 2006.

Interior has an annual budget of nearly \$10 billion. The one-third of its programs rated results not demonstrated represent 42 percent of its FY 2006 funding these include the Bureau of Land Management – Southern Nevada Land Sale, which was funded at \$735 million in FY 2006. Several Fish and Wildlife Service programs with relatively large budgets were rated results not demonstrated: National Wildlife Refuge System funded at \$391 million, Wildlife and Sport Fish Restoration \$791 million and Endangered Species, \$274 million.

HUD also has a high percentage of its budget rated ineffective at \$9 billion. This is due to two programs: Project Based Rental Assistance (\$5 billion) and the Community Development Block Grant (\$4.1 billion).

Defense is an agency with a large percent of high performing programs. Thirty one percent are rated effective, representing 43.6 percent of their budgetary authority in FY 2006. This includes large and established programs such as Military Force Management (\$109 billion), Navy Ship Operations (\$10 billion), Air Force Aircraft Operations (\$12 billion) and Army Land Force Operations (\$10 billion), Navy/Marine Corps Air Operations (\$6.5 billion) and Basic Skills and Advanced Training (\$4.7 billion).

¹³ The budget amounts given in PART for individual programs do not represent budget authority or outlays, but “funding levels” as reported by OMB. A program's “funding level” may include other kinds as spending such as fees, and offsetting collections. In order to calculate the percent of each agency's budget associated with a particular PART score, we take as our numerator the sum of all funding levels for programs in that agency with a particular score, and calculate it as a percentage of the agency's total budgetary authority as reported in the agency's annual financial statements. Due to this mismatch, some fractions may exceed 100%. budgetary authority as reported in the agency's annual financial statements. Due to this mismatch, some fractions may exceed 100%.

Table 8: PART ratings according to agency budgets for FY 2006¹⁴

	Results not Demonstrated	Ineffective	Adequate	Moderately Effective	Effective
Agriculture	4.2%	0.0%	24.4%	63.7%	7.8%
Commerce	2.8%	0.0%	38.1%	48.1%	10.9%
Defense	4.0%	0.0%	17.9%	34.5%	43.6%
Education	8.0%	1.8%	70.7%	17.5%	1.9%
Energy	1.1%	0.3%	36.1%	38.6%	23.9%
HHS	1.3%	0.3%	32.7%	60.5%	5.3%
DHS	21.6%	0.0%	34.4%	38.4%	5.6%
HUD	22.3%	25.3%	1.7%	47.0%	3.7%
Justice	6.3%	0.2%	41.9%	41.0%	10.5%
Labor	0.8%	14.9%	27.5%	53.4%	3.3%
State	1.5%	0%	40.7%	19.1%	38.7%
Interior	42.9%	0.0%	31.7%	19.7%	5.6%
Treasury	13.7%	1.0%	29.2%	35.9%	20.3%
Transportation	0.2%	2.0%	14.6%	72.7%	10.6%
Veterans'					
Affairs	51.0%	0.0%	46.2%	2.8%	0.0%
EPA	1.3%	5.1%	81.4%	12.3%	0.0%
NASA	0.0%	0.0%	44.0%	36.6%	19.4%
NSF	0.0%	0.0%	0.0%	0.0%	100.0%
SBA	0.0%	0.0%	5.1%	5.8%	89.1%
SSA	0.0%	0.0%	0.0%	100.0%	0.0%
GSA	25.0%	0.0%	13.9%	35.4%	25.6%
NRC	0.0%	0.0%	0.0%	41.7%	58.3%
USAID	2.1%	0.0%	63.7%	34.2%	0.0%
OPM	0.1%	0.0%	99.7%	0.2%	0.0%
USACE	58.0%	0.0%	8.4%	33.6%	0.0%
OTHER	36.1%	0.1%	4.2%	55.2%	4.3%

¹⁴ On <http://www.expectmore.gov>, OMB provides a breakdown of each agency's FY2008 proposed funding by PART rating. We use FY 2006 actual budgetary authority. This produces some discrepancies for some agencies.

Chart 4: Programs rated results not demonstrated by agency

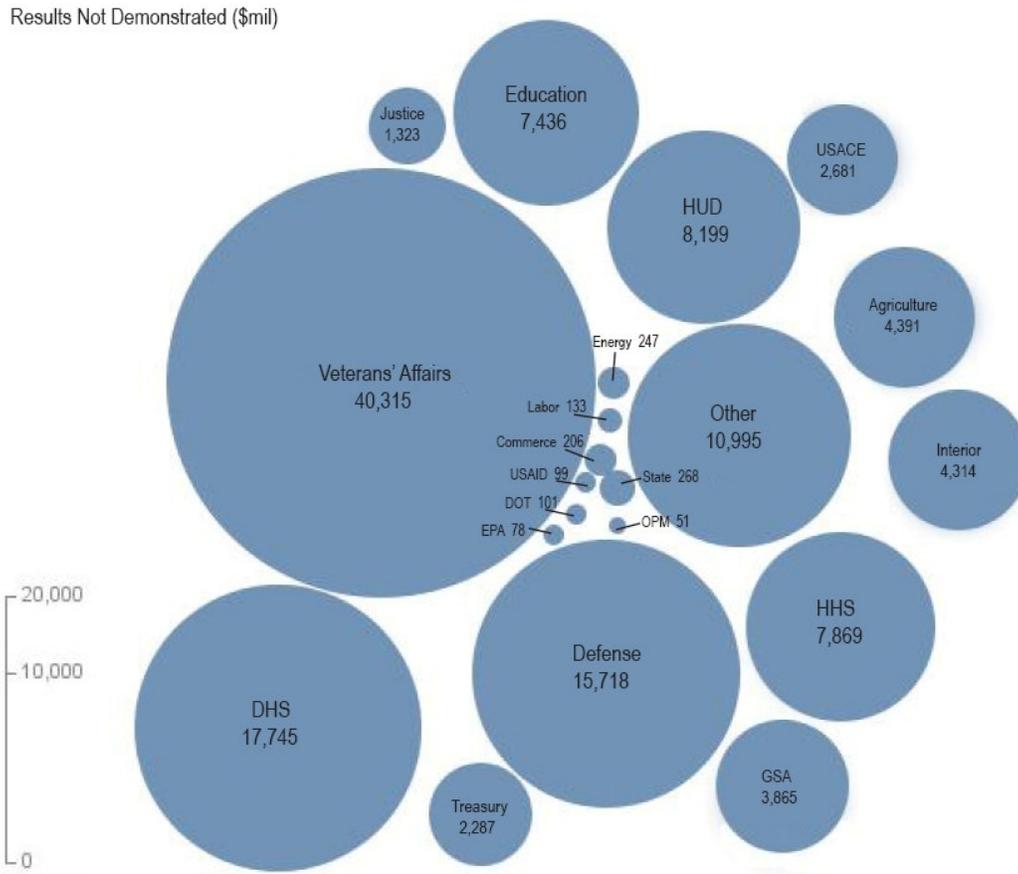


Chart 5: Programs rated ineffective by agency

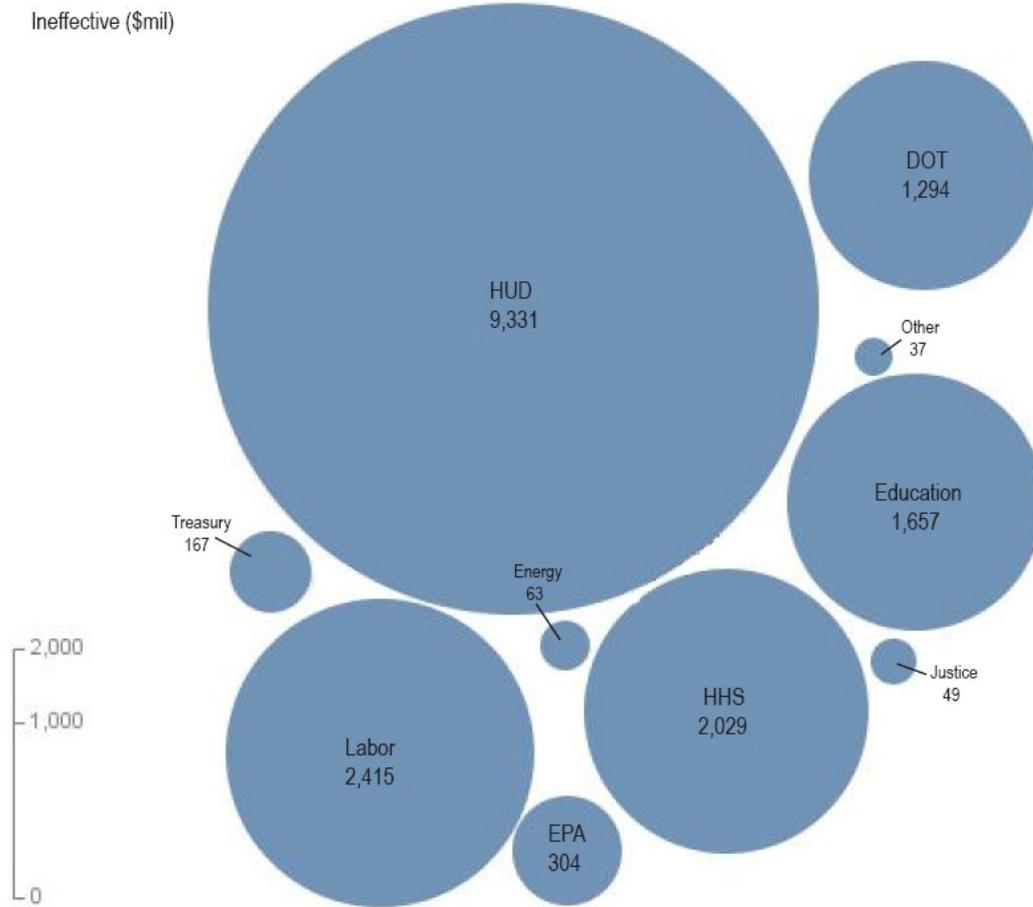


Chart 6: Programs rated adequate by agency

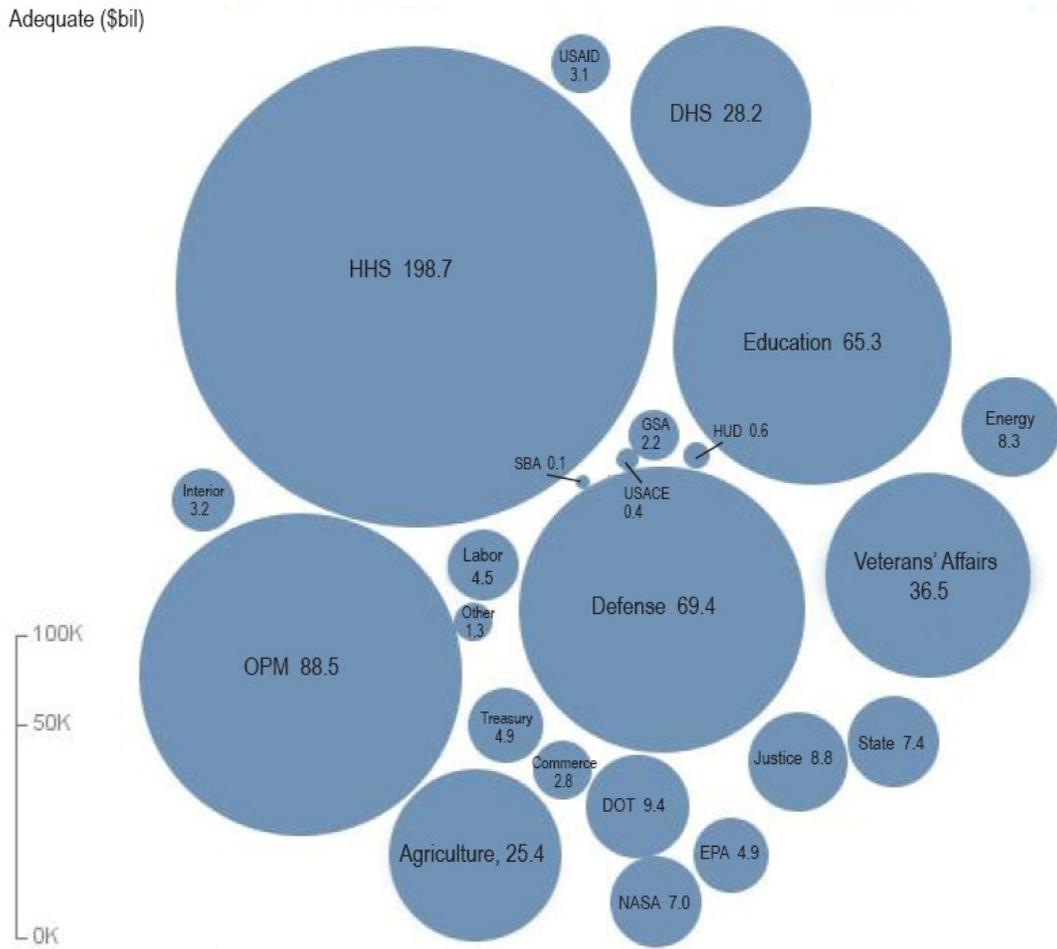


Chart 7: Programs rated moderately effective by agency

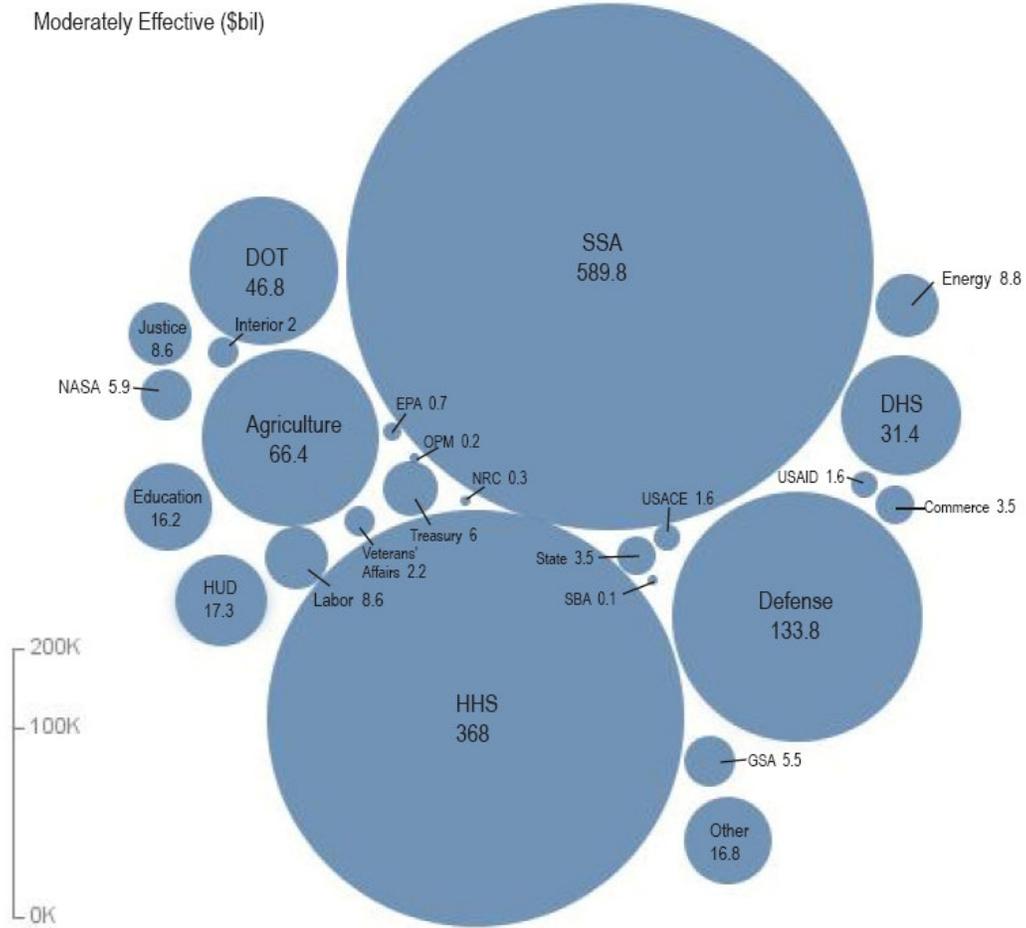
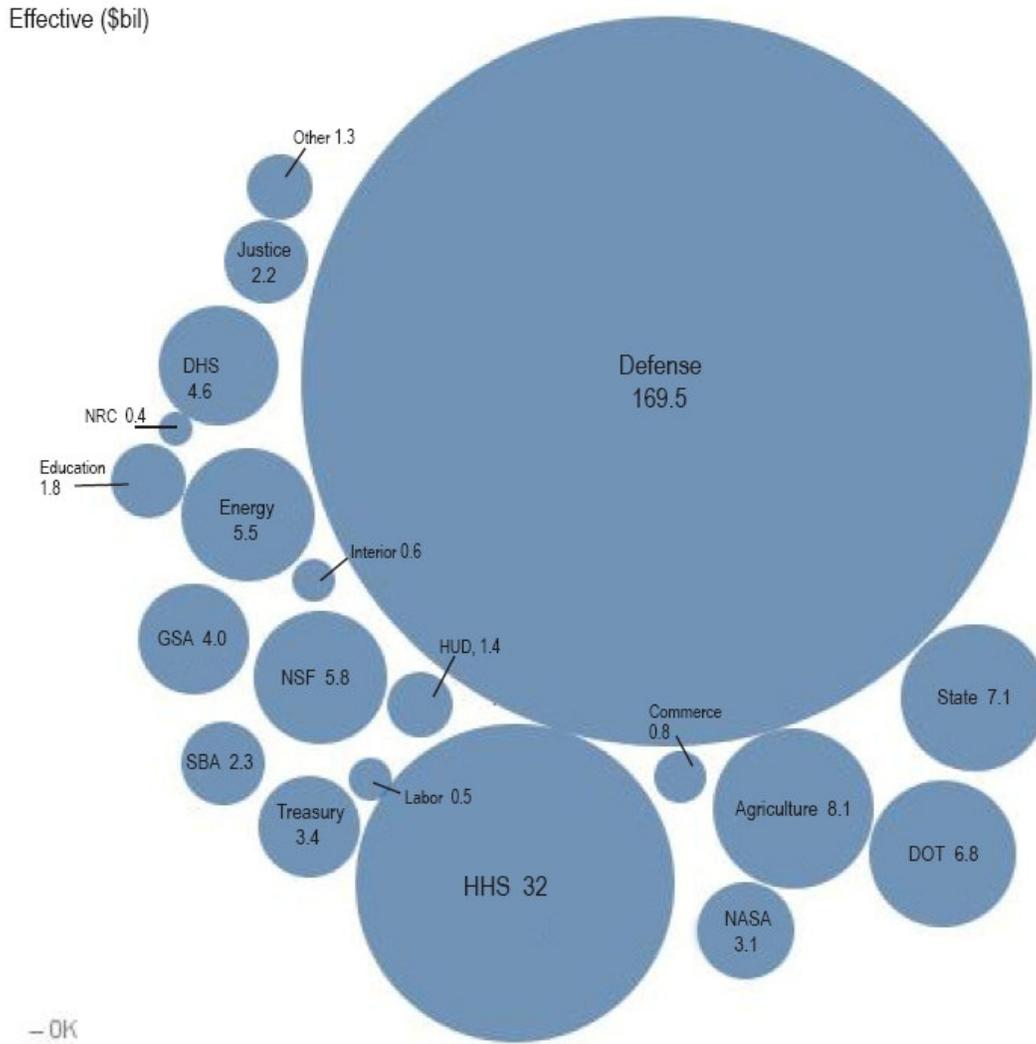


Chart 8: Programs rated effective by agency

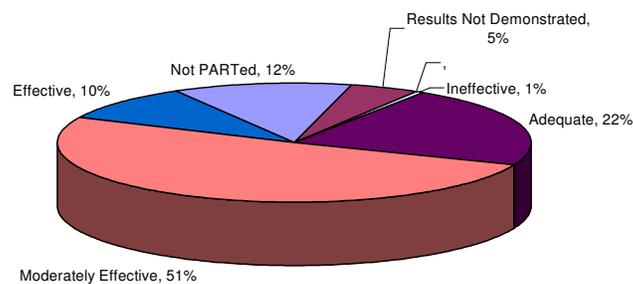


Percent of Budget represented by different PART ratings

In FY 2008, OMB completed its assessment of all programs in the federal budget. This allows us to see how much of the federal budget falls into certain ratings categories. Half of federal outlays in for FY 2006 are rated moderately effective. Moderately effective programs include two of the largest in the federal budget: Social Security and Medicare. Ineffective programs account for 1 percent of the federal budget (\$17 billion in funding levels) and results not demonstrated account for 5 percent of total outlays in FY 2006, representing \$128 billion in funding levels as reported by OMB.¹⁵

Chart 9

Percentage of FY06 Outlays by PART Rating



Presidential Funding Trends

The difference between the president's FY 2008 funding request and Congressional appropriations in FY 2007 for PARTed programs shows there is a tendency for the President to recommend funding increases for effective and moderately effective programs and decreases for ineffective and results not demonstrated programs. This

¹⁵ The pie chart shows 12 percent of the budget as not yet PARTed. OMB reports that 96 percent of the budget has been PARTed. This is due to a discrepancy in how these figures were calculated. We took the total amount of funding levels as reported by OMB in PART according to ratings category as our numerator. Summing all the funding levels leads to a total of \$2,333,141,000,000 using FY 2006 funding levels. Total outlays in FY 2006 amounted to \$2,655,435,000,000. We use this as our denominator. The difference between these figures represents \$322,294,000,000. Funding levels are not strictly comparable with outlays, thus our calculations are only proxies for the portion of the budget represented by PART ratings.

analysis does try to determine to what extent the PART scores are used in making these funding decisions, versus other political factors.

Several recent studies have examined how PART affects budget-making. In 2005, GAO analyzed the effect of PART scores on funding levels for discretionary programs for the first year of PARTed programs.¹⁶ They found for every one point increase in PART, the budget for all discretionary programs increased by 0.5 percent. GAO further divided programs into large, medium and small, based on the amount of their individual budgets. They found that PART had a significant effect on recommended funding levels for small-sized programs, but not for large and medium sized programs.

Gilmour and Lewis note that GAO's study is limited because it does not isolate the political factors that necessarily influence budget decisions.¹⁷ Using only FY 2004 data they attempt to isolate the effect of merit (as represented by PART) from political influences, measured as whether a program is housed in a "Democratic" or "Republican" agency. They make their determinations by grouping agencies according to general policy leanings and the platforms of each party.¹⁸ Additionally Gilmour and Lewis create dummy variables to indicate whether a program was created during a Democratic president, Democratic Congress, or unified government.

Looking at the first 234 programs evaluated by OMB, Gilmour and Lewis find that both PART scores and political support influenced funding levels. In particular, management scores were positively correlated with proposed budgets for programs in agencies characterized as Democratic, but not in other programs. They conclude that merit (PART scores) mattered more for programs in Democratic agencies, possibly because the administration insulates programs it favors from the consequences of evaluation. They conclude that since budgetary decisions operate in highly political environments, it is interesting that PART has any effect – indicting that merit does play a role. However, since the impact is limited to Democratic programs this suggests that political preferences still play a role.

In another article, Gilmour and Lewis investigate the role of PART and political factors on budget changes in FY 2005. They find, contrary to their previous findings, that programs in traditionally "Democratic" agencies are not more likely to receive budget increases or decreases than other programs. Building upon GAO's earlier study, they also find that PART scores have a larger effect on small and medium-sized programs, than for larger ones – an increase of 10 points in a PART score raises the program's budget from 1.3 to 13.5 percent.¹⁹

¹⁶ Government Accountability Office, *Performance Budgeting: Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, (January 2004) GAO-04-174.

¹⁷ John B. Gilmour and David E. Lewis, "Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores," *Public Administration Review* (September/October 2006): 742-752.

¹⁸ For instance, HHS, HUD, Labor and EPA are considered "Democratic". Since the Departments of Commerce, Education, and Energy were recommended for elimination by the Republican Congress in 1994, these agencies are also classified as "Democratic."

¹⁹ Gilmour and Lewis: "Assessing Performance Budgeting at OMB" (2005): 184.

The difference in results: in FY 2004, political content of programs mattered in funding levels, in FY 2005, they had no effect, may be due to how the variables were able to pick up on the political content of programs in different agencies. Overall, the findings of these studies suggest that PART does play a role in the President's funding decisions, and that though Democratic programs tended to be cut (indicating a possible political bias), these decisions were based to some degree, on merit.

Congressional appropriations to PARTed programs (FY06-FY07)

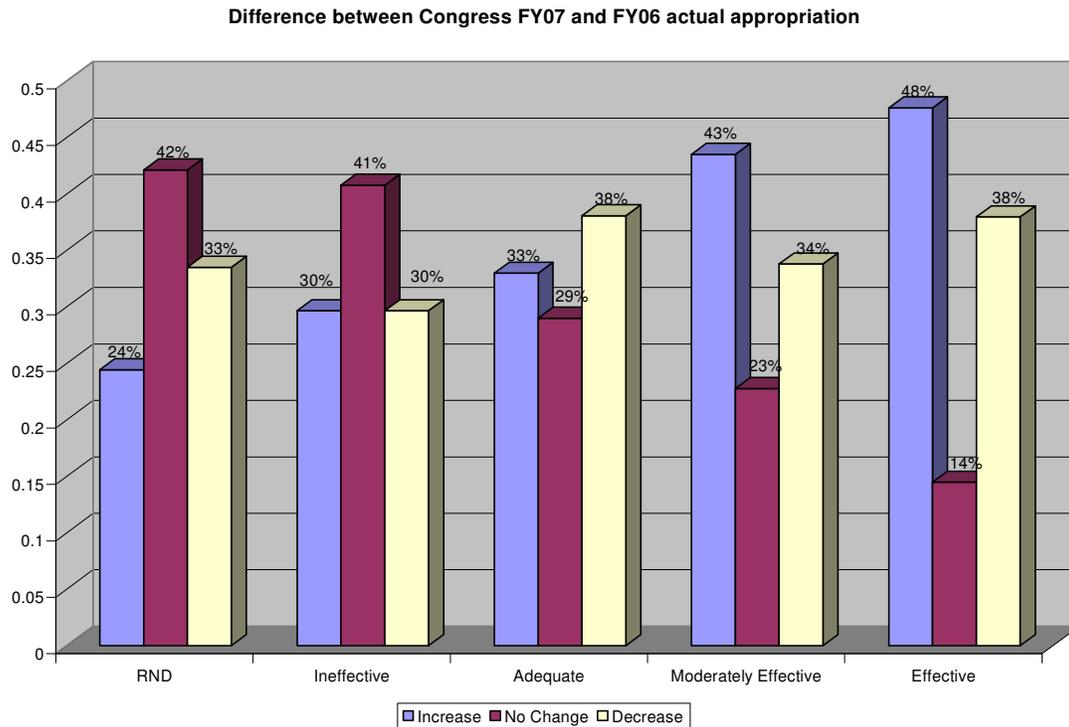
There is little evidence, in the fifth year of PART, that Congress uses PART scores in making funding decisions. Nevertheless, there is a tendency for Congress to award more budget increases to effective and moderately effective programs, and fewer increases to ineffective and results not demonstrated programs, though not to same degree as the administration. For instance, the administration recommends that 63 percent of effective programs receive an increase in FY 2008. Congress gave increases to 48 percent of effective programs in FY 2007. While 52 percent of ineffective programs are recommended for funding decreases by the administration in FY 2008 Congress awarded 30 percent of ineffective programs funding decreases.

This tendency may indicate that some PART ratings are somewhat in sync with Congressional views of particular programs – that is, PART is identifying programs that both the Administration and Congress view as deserving of budgetary increases, for any number of factors: performance, political or other considerations. Conversely, the causality may run in the other direction – programs that historically receive increases, tend to get higher ratings from the Administration.

The programs in this grouping include long-running government services such as the Mint, Bureau of the Census, and other well-established programs: Visa and Consular Services, New Currency Manufacturing. However, comparing the amounts recommended for by the President for individual programs rated effective and the amounts awarded by Congress shows a great deal of divergence and disagreement over funding levels between the Executive and Congress.

A greater proportion of ineffective and moderately effective programs are recommended for increases by the administration, and awarded increases by Congress. Yet this tendency does not hold as strongly for ineffective programs. The administration makes clear that ineffective programs tend not to receive funding increases (only 15 percent are recommended), while Congress increased budgets for 30 percent of ineffective programs.

Chart 10



The president’s *Major Savings and Reforms* report for FY 2008

For the third year in a row, the Bush Administration has issued a *Major Savings and Reforms* report. A supplemental to the president’s recommended budget, the document lists all the programs recommended for elimination, reduction, or reform. In FY 2008, a total of 141 programs are recommended for either reduction or termination. This is the same number of programs the administration recommended for cutting or elimination in FY 2007 – though, the lists contain a few differences. The Appendix includes the full list of programs for FY 2008.

In FY 2006, the president recommended a total of 154 programs for either cutting or elimination, and in FY 2007, 141 programs were recommended for budget cuts or elimination. Congress accepted 89 of the president’s 154 proposals – in whole or in part – representing \$6.5 billion in reduced spending.

Table 9: Proposed terminations for PARTed programs and funding amounts in FY 2008 Major Savings and Reforms

	Results not Demonstrated	Ineffective	Adequate	Moderately Effective	Effective	Total
Number of Programs	22	5	6	4	0	37
Dollar Amount Proposed for Termination (\$mil)	2962	311	202	2587	0	6062

Of the 91 recommended terminations, we were able to identify 37 programs as having been PARTed. Many of the other programs were either smaller items within larger programs, demonstration projects, or earmarks, and thus do not appear in PART. Of these 37 programs that had been PARTed 22 were results not demonstrated and five were ineffective. In total, these recommended terminations of PARTed programs represented \$6 billion funding.

In many cases, the PART assessment is cited as evidence for a recommended termination. Additionally other reasons are given – not a federal priority, duplication with other programs, or poor performance based on other evidence, such as independent assessments.

The administration recommends that 50 programs have their funding cut in FY 2008, for an overall reduction in spending of \$7 billion. We identified 34 of these programs as having been PARTed. Unlike program terminations, a good number of programs were moderately effective (11) and ten programs were rated adequate. Nine were rated results not demonstrated, and four were rated ineffective. These 34 programs represent \$4 billion in program reductions.

It is not possible to determine, in this analysis, the degree to which these various factors play a role in the decision to recommend a program’s termination or reduction in funding.

Table 10: Ratings for PARTed programs suggested for reduction in the FY08 budget

	Results not Demonstrated	Ineffective	Adequate	Moderately Effective	Effective	Total
Number of Programs	9	4	10	11	0	34
Dollar Amount Proposed for Reduction (\$mil)	781	2023	540	997	0	4341

How is PART used in budgetary decision-making?

PART is the Bush Administration’s attempt at performance budgeting – basing budgets on how well programs deliver results. As Gilmour and Lewis note, it is the most recent effort to resolve the “lack of a budgetary theory” that is, since governments operate outside the market and prices, on what basis does a policymaker decide how to allocate dollars among competing uses? This raises several important questions: a) is PART a reliable tool for measuring program performance, that is, do the ratings reflect reality? b) Even, where well-designed, how well do outcome measures capture program performance and can they be reliably employed to make policy judgments? And c) Are there any biases in how PART scores are assigned. That is, do programs favored by the administration receive higher scores? These questions are beyond the scope of this paper, but are receiving more attention from scholars reviewing PART for evidence of its effects and application.

In the *Major Savings and Reforms* report, the Administration notes that programs were recommended for funding reduction or elimination based on whether the program met the Nation’s priorities, constituted an appropriate and effective use of Federal dollars and produced the intended result. PART figures into some of these decisions in executive budget-making. This has advanced the idea that budget decisions should be based to some degree on performance information, and not only on political calculation or policy preferences of the administration.

However, as an initiative of the administration, PART has not been employed in Congressional budgets. This has lead to debate that in order for a tool such as PART to gain wider acceptance, program assessments are more appropriately conducted by an independent entity. This, it is argued, would remove any possible bias in how scores are assigned or applied.

III. Conclusion

PART is the first consistent and comprehensive attempt to adopt a form of performance budgeting in the federal government. Five years since its debut, PART has assessed nearly all government programs. PART's design and application raises many interesting questions in the public administration literature about how PART actually affects budgetary decisions. Evidence shows that it does play a role in the president's funding recommendations, though, not surprisingly, political preferences are also shown to figure into the budgetary calculation. Other questions that have been raised include: can PART – or any system of program evaluation – extract useful information about program performance for appropriators? Does PART advance the goals of GPRA or create its own 'paperwork requirements' creating compliance, but not necessarily better performance?

In this paper, we simply ask: how are agencies doing in PART? What patterns emerge among program types or agencies? Are there systematic differences in agencies or program types that account for varying degrees of performance?

Our main findings are that over the past five years some agencies have been high performers: State, Defense, Transportation and Energy. Several of these agencies have been actively pursued reassessment, and improved their ratings as a consequence. Other agencies, such as Education and Veterans Affairs have consistently performed poorly on PART. Education may be deliberately signaling that it cannot make its current portfolio of programs work, and wants them to be restructured or terminated.

The percent of programs rated results not demonstrated has dropped significantly. In the first year of part, 50 percent of the 234 programs assessed (117 programs) received this rating. In FY 2008, with 973 programs assessed, the percentage dropped to 21 percent (204 programs). The percent of programs receiving high ratings has grown over time. This may indicate that in later years, agencies were prepared to take the PART assessments based on their initial experience in 2004, or that OMB has been relatively lenient. It could also be that the GPRA requirements have created better reporting across agencies. Much of the information – strategic planning, purpose and design and management can feed into PART. As agencies improve their reporting, it may be that they also improve their PART scores.

How this information is applied in the President's budget cannot be determined by this analysis. We can observe general trends – effective programs tend to get increases, while ineffective programs tend to get decreases. The administration uses PART as one factor in its *Major Savings and Reforms* report, but also recommends programs for termination based on other criteria. Congress has not shown much interest in using PART to make budget decisions, mainly regarding it as an initiative of one administration. This analysis shows that funding trends tend to be aligned with PART scores in some cases (effective programs tend to get more increases, than ineffective programs), but this could be due to a variety of factors. Do programs that have historically received increases tend to be rated more favorably in PART? That is, does the previous budget affect the PART score? If

budgeting is incremental, then existing budgetary dynamics may in some cases drive the PART score.